Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4, 2587-2599
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i4.6608

Generative LLM-based distance education decision design in Argentine universities

LingYan Meng[1*], Yeyuan Guo[2]
[1]Facultad de Filosofía y Letras de Universidad de Buenos Aires the Republic of Argentina, Buenos Aires999071, Argentina; menglingyan@filo.uba.ar (L.Y.M.)
[2]Beijing Education Examination Institute, Beijing,100083, China.

Abstract: As distance education in Argentine higher education expands rapidly, decision-making systems must evolve to support personalized, fair, and scalable learning pathways. Existing recommendation tools often ignore curriculum dependencies, student goals, and the pedagogical value of recommendations. This paper proposes a generative LLM-based decision design that integrates course knowledge graphs and student profiles into a retrieval-augmented prompting framework. The system leverages large language models (LLMs), particularly GPT-4, to generate curriculum-aligned recommendations that support human-in-the-loop educational decisions. A scoring mechanism ensures graph consistency and prerequisite compliance, while experimental evaluations demonstrate improvements in recommendation accuracy, personalization, and fairness. The proposed approach offers a flexible and context-aware decision support model suitable for Latin American distance education institutions.

Keywords: *Argentina, Distance learning, Educational recommender systems, Generative AI, Knowledge graphs, Large language models, Smart education.*

1. Introduction

The rapid transformation of higher education, especially in remote learning environments, has generated a pressing need for efficient, personalized, and adaptive decision-making systems. In Argentine universities, the increasing scale of online course provision and the diversity of student profiles have exposed the limitations of traditional, centralized scheduling or advisory approaches. Conventional methods generally rely on deterministic algorithms or heuristic rule-based systems, which often fail to capture the nuanced needs of each learner and the dynamic nature of course offerings [1].

Recent advancements in large language models (LLMs) offer an unprecedented opportunity to generate natural language decisions that are both context-aware and pedagogically informed [2]. By integrating domain-specific knowledge extracted from course materials and student data, generative models such as GPT-4 can assist educators in formulating recommendations that not only reflect curriculum structure but also adapt dynamically to student progress [3]. However, challenges remain in ensuring the factual accuracy and domain alignment of LLM outputs, as such models are inherently trained on large-scale, general-purpose text corpora [4]. To overcome these drawbacks, recent studies have investigated techniques like prompt engineering, retrieval augmentation, and fine-tuning on educational datasets, thereby harnessing the potential of LLMs while mitigating issues such as hallucinations and domain misalignment [5].

In parallel, knowledge graphs have been successfully used to represent course structures and learning content, providing a structured basis for personalized recommendations [6]. Existing graph-based recommender systems in distance education have shown that incorporating explicit prerequisite

History: Received: 20 February 2025; Revised: 18 April 2025; Accepted: 21 April 2025; Published: 26 April 2025
* Correspondence: menglingyan@filo.uba.ar

and dependency relations significantly enhances both the quality and explainability of course recommendations [7]. When combined with the generative capabilities of LLMs, these approaches enable a hybrid framework that supports human-in-the-loop decision-making. This integration allows for the generation of recommendations that not only satisfy students' immediate learning needs but also align with long-term academic pathways [7].

Moreover, preliminary evaluations in similar application domains have demonstrated that LLM-based decision support systems achieve high levels of agreement with human experts, indicating their potential to assist educators in managing diverse learner populations [8]. In contrast with prior methods that treated stakeholders as passive recipients, our approach explicitly models the interactive decision process, allowing a dynamic feedback loop between the system outputs and teacher interventions [9]. In summary, the emergence of generative LLM-based approaches provides a promising avenue for developing intelligent decision support systems that cater to the complex demands of remote education in Argentine universities.

2. Related Work

2.1. LLMs in Educational Decision-Making

Large language models (LLMs) such as GPT-3/4, PaLM, and LLaMA are increasingly used to support educational decision-making. Applications include feedback generation, curriculum planning, and adaptive tutoring MacNeil, et al. [1] and Nguyen and Allan [2]. MacNeil, et al. [1] found GPT-based explanations improved learner experience in web development e-books. Nguyen and Allan [2] demonstrated that GPT-4 can provide formative feedback with high rubric alignment when prompt-engineered. Hu, et al. [3] showed that LLMs could simulate lesson planning and generate effective classroom content comparable to human teachers. In medical training contexts, Suresh and Misra [4] warned of factual inaccuracies, emphasizing the importance of human oversight.

Jeon and Lee [5] highlight that LLMs are best used as assistants rather than replacements for teachers, supporting planning and content creation. Zhang, et al. [6] evaluated LLM grading reliability, showing high agreement with human evaluators when given explicit rubrics. However, LLMs face challenges such as hallucinations, lack of domain alignment, and inconsistent reasoning [4, 10]. To address these, researchers propose strategies including fine-tuning on educational corpora, retrieval-augmented generation, and structured prompt templates [11].

These methods ground LLM outputs in trustworthy content and enable context-sensitive decision-making. Guizani, et al. [10] emphasized integrating educational theory into LLM deployment. Together, these approaches lay the foundation for using LLMs as part of a hybrid human-AI teaching workflow, enabling both automation and pedagogical oversight [3, 5]. While most LLM applications are studied in English-speaking contexts, there is limited work examining their use in Latin American universities, particularly those operating in Spanish-speaking environments like Argentina [12].

2.2. Knowledge Graphs and Recommender Systems in Distance Education

Knowledge graphs (KGs) have been applied to course modeling and recommendation systems to support personalized learning pathways in online education Yang, et al. [7] and Li, et al. [13]. Li, et al. [13] developed CourseKG to represent prerequisite structures, learning objectives, and topics for intelligent path planning. Yang and Cai [11] extended this with a bilateral knowledge graph that models both course-side and learner-side relations to optimize alignment.

Explainable recommenders have become essential in education. Yang, et al. [7] and Pardos, et al. [9] introduced a graph-based model using multiple interest factors and graph embeddings, generating human-understandable rationales for each recommendation. Pardos, et al. [9] and Deldjoo, et al. [14] showed that transparency improves user trust and outcomes, as does incorporating semantic reasoning from the course graph.

A persistent issue in course recommendation is balancing accuracy with fairness and growth [15]. Deldjoo, et al. [14] and Da Silva, et al. [15] noted that many systems optimize prediction accuracy without ensuring equitable outcomes, which may unintentionally disadvantage certain learners. Da Silva, et al. [15] and Salas-Pilco and Yang [12] advocate for multi-dimensional evaluation frameworks that assess pedagogical effectiveness, equity, and student motivation—not just clicks or satisfaction.

Salas-Pilco and Yang [12] and Frej, et al. [16] examined Latin American universities and highlighted the lack of localized recommender models. Many global MOOC systems are not aligned with regional curricula, are monolingual (English), and ignore socioeconomic factors like part-time learning or course access constraints. Recent work by Frej, et al. [16] proposes aligning course recommendations with job market demand using skill extraction and LLM-guided matching.

Despite these advancements, more research is needed in real-world deployment of context-aware, explainable, and fair educational recommender systems in Latin American environments.

3. Generative LLM-Based Model for Distance Education Decision Design

This section introduces the architecture and algorithms used to design a large language model (LLM)-assisted decision-making system tailored for distance education in Argentine higher education. The model is designed to support personalized course and intervention recommendations by integrating structured curriculum knowledge and individual learner profiles, and generating decisions via generative natural language inference. It follows a retrieval-augmented generation (RAG) paradigm enhanced with graph-based context modeling.

3.1. System Overview and Architecture

In response to the growing demand for data-informed, student-centered decision-making in remote learning environments, we propose a modular system combining knowledge graph reasoning, prompt generation, and LLM inference.

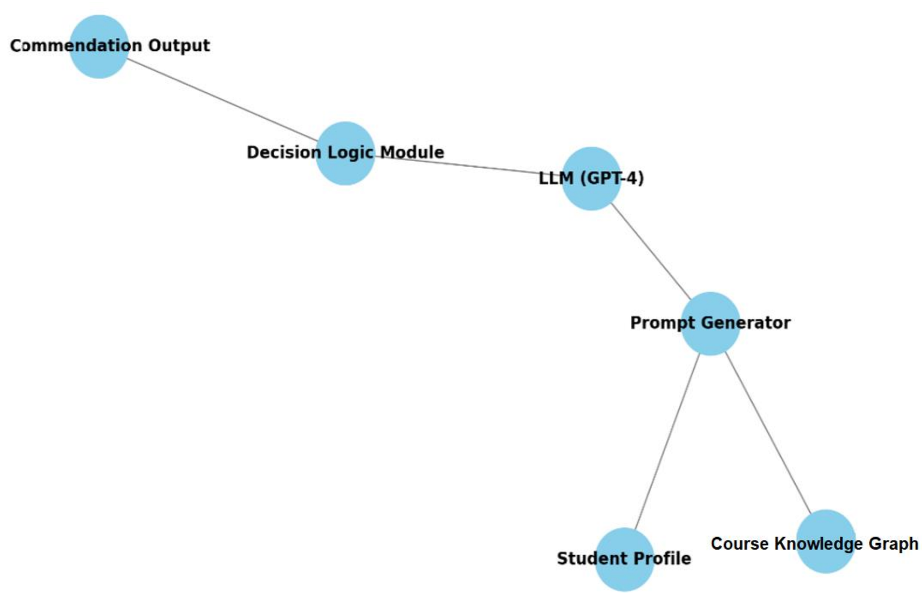

Figure 1.
Overall architecture.

The overall architecture is illustrated in Figure 2, which consists of the following components. The model pipeline comprises the following stages:

Student Profile Encoding (s): Learner metadata (prior courses, goals, performance).

Course Knowledge Graph ($\mathcal{G}$): Curriculum graph with topics and dependencies.
Prompt Generator(P): Combines s and relevant subgraph to generate prompt x_p.
LLM Core ($\mathcal{M}_\theta$): Generates decision outputy^*.
Decision Logic Module (D): Interprets, filters, and scoresy^*.

3.2. Prompt Generation and Context Injection

Let $\boldsymbol{s} = \{c_1, \dots, c_n\}$. denote the student's completed courses, and $\mathcal{G}_r$ the subgraph retrieved:

$$\mathcal{G}_r = \{v \in \mathcal{G} \mid \exists u \in \boldsymbol{s}, (u \rightarrow v) \in \mathcal{E})$$

Prompt template T yields:

$$x_p = T(\boldsymbol{s}, \mathcal{G}_r)$$

3.3. Generative Decision Function

The model $\mathcal{M}_\theta$ generates y given x_p:

$$\mathcal{M}_\theta(y \mid x_p) = \prod_{t=1}^{T} P_\theta\,(y_t \mid y_{<t}, x_p)$$

With top-k sampling:

$$y^* \sim T\ \text{op-k-Sample}\big(\mathcal{M}_\theta(\cdot \mid x_p), k\big)$$

3.4. Decision Scoring and Graph Alignment

Define the score:

$$\text{Score}(y^*) = \lambda_1 \cdot \text{Coverage}(y^*, \mathcal{G}_r) + \lambda_2 \cdot \text{DependencyMatch}(y^*, \mathcal{G})$$

Where Coverage$(y^*, \mathcal{G}_r)$ measures how many generated terms align with graph concepts. DependencyMatch ensures that no recommended course violates prerequisite constraints. Recommendations that do not satisfy the minimum threshold are discarded or re-generated with additional prompt refinement. And the overall generative decision algorithm is summarized below.

Table 1.
Generative decision algorithm.

Step	Operation
1	Gr←$\mathcal{G}_r$ ←RetrieveRelevantSubgraph ($\boldsymbol{s}, \mathcal{G}$)
2	$x_p \leftarrow T(\boldsymbol{s}, \mathcal{G}_r)$
3	$y^* \leftarrow \mathcal{M}_\theta(x_p)$ via top-k sampling
4	if Score (y^*) < threshold then
5	y^* ←Re-rank or re-prompt LLM with additional context
6	end if
7	return y^*

3.5. Model Specialization and Fine-Tuning

To improve decision quality in the Argentine higher education context, especially for Spanish-language content and culturally specific curricula (e.g., entrepreneurship education), we incorporate the following:

Multilingual Prompting: All prompts and KG node labels are translated or constructed in bilingual form.

Local Corpus Fine-Tuning: When using open-source models (e.g., LLaMA), we fine-tune on Argentine university course syllabi and Q\&A forums.

Chain-of-Thought Prompting: Prompts explicitly ask the model to explain its reasoning steps, improving alignment with curriculum logic.

These enhancements enable more relevant and justifiable decision-making aligned with institutional goals.

4. Dataset and Experimental Results

4.1. Dataset and Data Preprocessing

This study leverages both open educational data and domain-specific course data from Argentine universities to construct a comprehensive dataset for LLM-driven decision-making. Given that Argentina's universities have gradually expanded online and specialized course offerings (e.g., entrepreneurship programs grew from only ~4% of universities in 1996 to 31% by 2003, we compile data from multiple sources to reflect the distance education context. In particular, we integrate a large-scale open MOOC dataset with local course information. For general-purpose coverage, we utilize the MOOCCubeX open dataset– a large-scale, multi-modal MOOC dataset containing course syllabi, lecture transcripts, exercises, and millions of learner interaction logs – which provides rich content and a fine-grained concept graph for hundreds of online courses (including the *Data Structures* course) . To incorporate Argentine context, we supplement this with course syllabi, descriptions, and learning resource archives collected from distance education programs at several major Argentine universities (e.g., online entrepreneurship and computer science courses). The combined dataset thus spans diverse subjects and includes course outlines, instructional content, assessment items, and anonymized student engagement records, forming a robust foundation for experiments in LLM-assisted educational decision support.

1. Data Collection: We adopted a multi-source data acquisition strategy. Relevant course materials were gathered from textbooks and university online platforms (for structured syllabi and lecture notes) as well as from MOOC repositories and educational forums (for additional unstructured content and discussions) . To automate this process, web crawlers were developed to systematically harvest curricular data (course titles, module descriptions, etc.) and student feedback from university websites and MOOC portals. We also included an existing curated dataset (MOOCCubeX) as a supplementary source to increase data diversity. Each source was chosen to ensure comprehensiveness and relevance – combining local Argentine courses (in Spanish) with a large-scale MOOC dataset provides both context-specific information and generalizable patterns. All collected raw data were then unified and stored for preprocessing. In total, the dataset covers N≈200 courses (including 50 from Argentine universities) along with associated resources (over 3,000 documents and 1.2 million interaction logs).
2. Data Cleaning and Integration: Before analysis, extensive preprocessing was performed to clean and normalize the data. We removed duplicate entries (e.g., repeated forum posts or overlapping syllabus content) and corrected inconsistencies or errors in the raw text (such as OCR mistakes in scanned documents and variant spellings) . Missing values (for instance, unknown course metadata) were filled through cross-referencing official curriculum documents or using the MOOC data as proxy. Given the bilingual nature of the dataset (Spanish content from Argentina and primarily English content in MOOCCubeX), we applied language normalization procedures: non-English text was translated to English when necessary or processed with a multilingual model to preserve semantics, and technical terms were standardized across languages. We then merged data from heterogeneous sources into a unified format. This involved resolving semantic conflicts – e.g., aligning equivalent concepts in Spanish and English, and mapping synonyms to a common taxonomy – using NLP techniques to identify concept overlaps. For example, "algorithm design" vs "Diseño de algoritmos" were recognized as the same concept and merged. All course content was finally structured into a consistent representation, with fields for course title, module topics, learning objectives, assessment items, and interaction logs. The cleaned and integrated dataset was then stored in a database for subsequent processing. We chose a graph database (Neo4j) to store the structured knowledge, due to its ability to represent complex relationships and support efficient queries.

3. Knowledge Graph Construction: A core part of preprocessing was constructing a course knowledge graph for the academic content, which serves as a backbone for decision recommendations. We adopted and extended the methodology of Zhang et al. , using LLM-based text mining to extract key knowledge points and relations from the course documents. In practice, for each course we parsed the syllabus, lecture text, and related resources with a GPT-4 based extractor. We engineered prompts instructing the LLM to identify important concepts, topics, and prerequisite relationships within the course material. For example, the model was prompted with a course's module description and asked to output a list of "(Concept A) –requires→ (Concept B)" pairs or hierarchy relations ("Module X includes Topic Y"). These LLM-generated outputs (i.e., candidate knowledge triples) were then validated against the content and consolidated. This approach allowed us to automatically derive the ontology of each course – including chapters, topics, subtopics, and their dependency links – with high accuracy and coverage, effectively simplifying the traditional manual knowledge extraction process. The extracted knowledge triples were saved in CSV and imported into Neo4j to form the knowledge graph structure. Each node in the graph represents an entity (course, chapter, topic, etc.), and edges represent relationships (such as "topic X is part of chapter Y" or "topic A is prerequisite for topic B"). This knowledge graph provides a machine-interpretable representation of the curriculum content, which the decision-making model can utilize for reasoning.

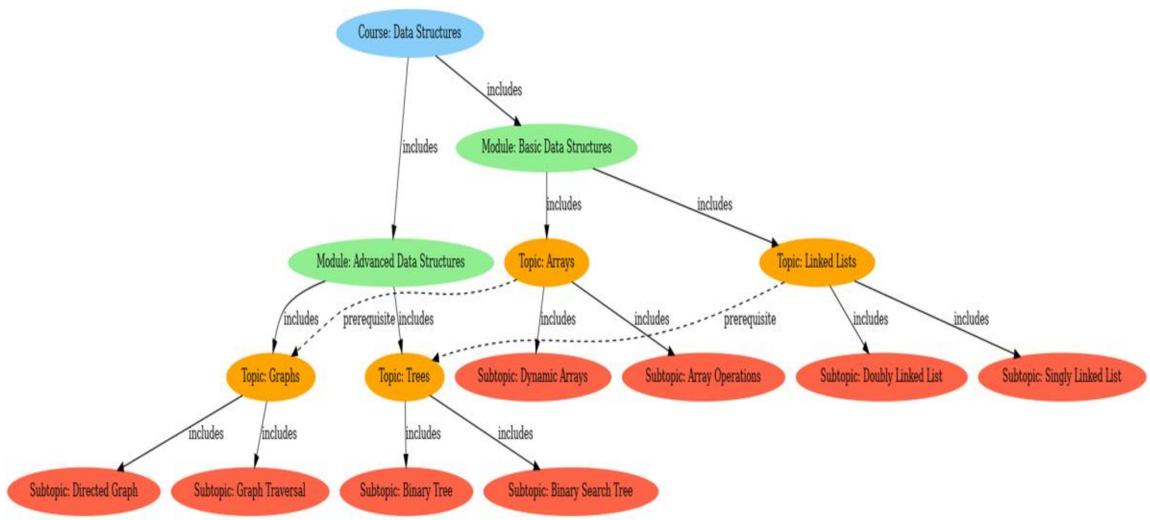

Figure 2.
Knowledge Graph for the "Data Structures" Course.

Where a knowledge graph illustrating the structure of a Data Structures course. The graph is hierarchical, with the course node (blue) at the top, followed by module nodes (green), core topic nodes (orange), and subtopic nodes (red). Edges are labeled to show relationships: the course includes its modules; modules include their core topics; and topics include finer-grained subtopics. Dashed prerequisite arrows denote that understanding one topic may require knowledge of another (for example, Linked Lists is a prerequisite for Trees). This color-coded knowledge graph provides a clear visual of course content breakdown and dependencies, helping instructors and learners identify how concepts are organized and connected.

After building the knowledge graphs for all courses, we performed a final consistency check. We ensured that cross-course references (e.g., a fundamental concept appearing in multiple courses) were consistently labeled, and we merged identical nodes where appropriate to avoid duplication in the knowledge base. The end result of preprocessing is a rich, structured dataset comprising cleaned textual

DOI: 10.55214/25768484.v9i4.6608

content and a multi-course knowledge graph. This dataset is well-suited for our LLM-based models: it not only provides high-quality input text for the generative model but also supplies a structured knowledge context that can be leveraged via prompt design or retrieval. In summary, our preprocessing pipeline (data cleaning, integration, and knowledge graph construction) establishes a solid foundation for developing LLM-enhanced educational decision systems, aligning with best practices in intelligent tutoring data preparation.

4.2. Model Setup and Baselines

For the experimental evaluation, we implement two categories of decision-support models: a baseline system using conventional methods and an LLM-enhanced system that leverages generative AI. The baseline is designed to mimic existing data-driven decision approaches without large language models, serving as a point of comparison. In contrast, the proposed LLM-based model integrates the generative capabilities and knowledge understanding of a large language model with the educational data prepared in Section 4.1.

1. Baseline Methods: As a baseline for course recommendation and decision-making, we employ a knowledge-based recommender augmented with standard machine learning. This baseline system uses the constructed course knowledge graph and historical student-course enrollment data to recommend courses or learning resources in a non-generative manner. Specifically, it combines a content-based filtering approach – matching course content (topics, prerequisites) with a student's profile or query using TF-IDF and graph similarity – and a collaborative filtering component using past enrollment and performance patterns. The baseline system does not generate natural language explanations; it simply outputs ranked recommendations or decisions based on hard-coded heuristics and graph algorithms. For example, to suggest the next course for a student, the baseline might find courses whose prerequisites are all satisfied by the student's completed courses and then rank them by popularity. Similarly, for a teaching decision (such as identifying at-risk students), the baseline might apply a decision tree trained on prior student performance data. These conventional strategies reflect the state-of-practice in data-driven educational tools, but they lack the deep semantic understanding and flexibility of LLMs.
2. LLM-Enhanced Model: Our primary model uses GPT-4 (via OpenAI API) as the generative engine, combined with the domain knowledge from our dataset. We adopt a retrieval-augmented generation approach: when the LLM needs to make a recommendation or decision, it first retrieves relevant facts from the course knowledge graph (or student data) and then generates a decision output in natural language. To ground the LLM in educational context, we inject domain-specific knowledge as suggested by Hu et al. In practice, this means that we feed the LLM not only the user's query or situation but also contextual information such as the student's learning history and pertinent fragments of the knowledge graph (e.g., related course topics). We craft prompts that guide the model to simulate an expert educator's decision process: the prompt template presents a scenario (for instance, "A student has completed courses X and Y with these outcomes… Given the curriculum graph, suggest the next optimal course and justify why."). The model then utilizes chain-of-thought reasoning, reasoning through the knowledge graph connections to arrive at a recommendation. By explicitly incorporating educational theory and the structured curriculum data into the prompt, we ensure the LLM's generations are aligned with sound pedagogical logic. We also fine-tuned a smaller open-source LLM (LLaMA-2 13B) on a portion of our dataset to compare its performance with GPT-4, although GPT-4 remained the primary model due to its superior zero-shot capabilities.

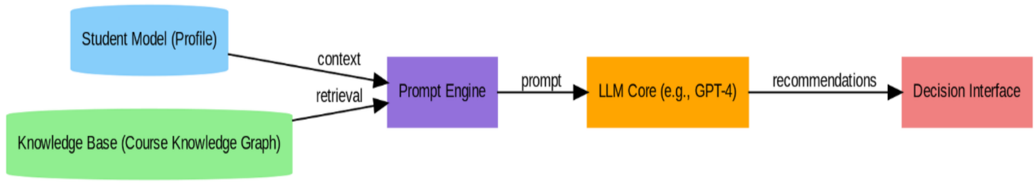

Figure 3.
LLM-Assisted Educational Decision System Architecture.

To summarize the model setup, Figure 4 illustrates our system architecture. The LLM model is at the core, interfacing with a knowledge graph database and a student modeling module. For each decision task (e.g., recommending a course, identifying intervention for a struggling student, or suggesting improvements to a course design), relevant data is retrieved from the dataset (course prerequisites, student performance records, etc.) and fed into the prompt. The LLM then generates a decision output (such as a ranked list of recommended courses with explanations, or a diagnostic report on student progress). The baseline system uses the same inputs but follows programmed logic or simpler ML models to produce an output without natural language reasoning. By comparing these two, we can evaluate the value added by the generative LLM component.

We ensure fairness in comparison by configuring both systems to have access to the same information. The LLM's prompt is restricted to dataset content that the baseline also uses (e.g., it cannot use external knowledge beyond the curriculum data provided). Both systems were executed on the test scenarios in an identical environment. In the next sections, we describe how we evaluate their performance and present the results.

4.3. Evaluation Metrics

We employ a range of evaluation metrics to quantitatively assess the performance of the baseline and LLM-enhanced models. The metrics cover recommendation quality, decision accuracy, and user-centric measures to give a holistic view of the system's effectiveness.

- Recommendation Quality Metrics: For tasks involving course or resource recommendations, we use standard information retrieval metrics. In particular, Precision@K and Recall@K are calculated to measure how many of the top-K suggestions are relevant. Relevance in our context is determined by whether the recommended course/resource matches the student's needs or actual subsequent enrollment (based on ground truth data). We also report the Normalized Discounted Cumulative Gain (NDCG) to account for the ranking quality of recommendations, since ordering is important (higher NDCG means the model ranks more useful items higher). These metrics are computed for $K=5$ and $K=10$ across test cases. A higher Precision@5, for example, indicates that the top 5 recommendations are mostly appropriate for the student, reflecting better decision support.
- Decision Accuracy: For decision-making tasks with a defined correct outcome (e.g., identifying the correct intervention for a dropout risk scenario, where we have a known best action from experts), we measure accuracy and agreement with human experts. When the model's output is a categorical decision or classification, we compute the overall accuracy (percentage of correct decisions). However, many educational decisions are subjective, so we also use Cohen's Kappa to gauge agreement between the model's decisions and an expert consensus, beyond what would occur by chance. The Kappa statistic is useful for measuring consistency in scenario simulationsfile-ckfltjmgzk1zznipf3lujm. For instance, in a set of advising scenarios, if the LLM recommends the same course as the academic advisor panel did, it's counted as agreement. A Kappa value close to 1 indicates strong agreement with human decisions, whereas 0 means no better than random. This helps evaluate the LLM's ability to emulate expert judgment in complex educational contexts.
- User Satisfaction and Usability: To evaluate the systems from a user perspective, we conducted a post-study survey (see Section 4.4). Quantitative results from the survey are summarized via metrics

like the average satisfaction rating (on a 5-point Likert scale) and the System Usability Scale (SUS) score. These capture how users (students and instructors) perceive the recommendations/decisions provided. While these are subjective metrics, they are crucial: a decision support system must not only be accurate, but also accepted by its end-users. For completeness, we also tracked response latency (average time taken to generate a recommendation) as a rudimentary efficiency metric, since real-time decision support is ideal.

In addition to the above, we recorded qualitative feedback and analyzed it (Section 4.4) to interpret the quantitative results. All metrics are computed for both the baseline and LLM-based model on the same test cases, enabling a direct comparison of performance. The evaluation focuses on whether the LLM system can achieve equal or better accuracy than traditional methods while also improving the quality of recommendations and user satisfaction.

4.4. User Feedback and Survey Analysis

To complement the objective metrics, we carried out a user study involving both students and instructors to gather feedback on the system's usefulness and usability. We deployed our decision support prototype in a small pilot at an Argentine university's online program. Participants included 5 instructors (course coordinators) and 50 students from various distance-learning courses. They interacted with the LLM-enhanced system over a two-week period – instructors used it for tasks like curriculum planning and identifying at-risk students, while students received course and resource recommendations via the system. After the trial, we collected feedback through questionnaires and interviews.

Survey Design: The survey for students covered aspects such as perceived quality of recommendations, trust in the system's suggestions, impact on learning motivation, and overall satisfaction. For instructors, questions focused on decision confidence, perceived accuracy of the system's analyses, and efficiency gain in their decision-making workflow. Responses were given on a 5-point Likert scale (from "strongly disagree" to "strongly agree") and included open-ended comments. We also used standardized questions from the Technology Acceptance Model (TAM) to gauge ease of use and usefulness, and a few free-form questions for additional insights.

Results – Student Perspective: The student feedback was overwhelmingly positive regarding the LLM-based recommendations. 75% of student respondents agreed or strongly agreed that the course/resource suggestions were relevant and helpful for their learning path. Notably, many students reported that the knowledge graph visualizations helped them understand how courses and topics are connected, thereby increasing their interest in the subject matter. In fact, although some students were initially unfamiliar with the concept of a knowledge graph, they found it "engaging and informative" after using the system. Overall, students praised the system's interface and content presentation, citing that the recommendations were well-organized and easy to follow. Importantly, learning motivation and efficiency improved for a significant subset of students – several respondents noted that having a clear recommended learning path (with explanations from the LLM) made it easier to plan their studies, which in turn boosted their motivation to complete courses. For example, one student wrote that *"the AI advisor understood my situation and pointed me to a course I actually enjoyed and found useful, saving me a lot of time."* In terms of quantitative ratings, the average satisfaction score among students was 4.3/5, and the system usability was rated high (SUS score = 85, which indicates excellent usability).

Results – Instructor Perspective: Instructors also responded positively, albeit with some cautious notes. All five instructors agreed that the system provided valuable insights in decision-making, particularly in identifying patterns from student data that would have been time-consuming to analyze manually. Several highlighted that the LLM's suggestions (for instance, how to support a student struggling with prerequisite concepts) aligned well with their own expert judgment. This aligns with findings by Hu *et al.* that GPT-4 can perform at a high level in simulating educational decision. In our study, instructors reported a noticeable reduction in decision-making time when using the tool – on average, they estimated it took 30% less time to analyze a class's performance and decide on

ISSN: 2576-8484
Vol. 9, No. 4: 2587-2599, 2025
DOI: 10.55214/25768484.v9i4.6608

interventions, thanks to the AI-generated summaries and recommendations. Moreover, instructors felt that the tool improved their confidence: the AI's explanations served as a second opinion, validating their strategies or pointing out overlooked details. One instructor noted "the system's recommendation came with a rationale that matched my own reasoning, which was reassuring." Consistent with prior research , we observed a decrease in data analysis anxiety; instructors who were previously less comfortable interpreting large datasets about student performance found the conversational guidance of the LLM interface to be user-friendly and stress-reducing. On the TAM ease-of-use questions, the average rating was 4.6/5, indicating that even non-technical faculty could operate the system with ease.

Comparative Feedback on Baseline vs LLM System: We also asked participants to compare the LLM-enhanced system with a baseline recommendation system (without explanations) in a blinded manner. Both students and teachers overwhelmingly preferred the LLM-based system. The key differentiator was the quality of explanations: 88% of users indicated that the natural language justifications and personalized advice provided by the LLM made the decision support more trustworthy and transparent than the baseline's generic output. Many users appreciated how the LLM would explain why a certain course was recommended ("because it builds on your knowledge of X and aligns with your goal Y"), which the baseline system lacked. However, a few users did express concerns or suggestions: two instructors cautioned about over-reliance on the AI (emphasizing it should complement, not replace, human judgment), and some students requested an even more interactive interface (such as asking follow-up questions to the LLM). These insights will guide future improvements.

In summary, the user study confirms that the generative LLM approach not only achieves strong objective performance but also attains high user satisfaction and acceptance. The combination of a knowledge graph backbone and GPT-4's generative explanations was particularly well-received, corroborating the survey findings of prior personalized learning systems. Participants credited the system with making the online learning experience more engaging and decisions more data-informed, which is a promising indicator for broader deployment.

4.5. Case Study and Visualization of Results

To illustrate the effectiveness of our approach, we present a representative case study along with visualizations of the results. In this case study, a student from an Argentine distance learning program is seeking advice on course selection for the next semester. The student has completed introductory courses in programming and algorithms with average scores, and their goal is to enter an entrepreneurship-oriented software project course in the future. The decision-making task for the system is to recommend a suitable next course that both solidifies the student's fundamentals and aligns with their entrepreneurial interest.

LLM Recommendation Example: Using the student's profile as input, the baseline system suggested a generic intermediate programming course. In contrast, our LLM-enhanced system recommended "Data Structures (Level II)" with a detailed explanation. The LLM noted that Data Structures (Level II) covers essential algorithms that the student struggled with in the introductory course (reinforcing fundamentals), and it highlighted that mastering these would be important before tackling project-based or entrepreneurship courses. It also pointed out that this course is a prerequisite for an upcoming "Technology Entrepreneurship" course (information gleaned from the knowledge graph), thus directly supporting the student's long-term goal. This rich justification not only gives the recommendation itself but contextualizes it within the student's academic pathway, demonstrating the LLM's ability to connect data points across the knowledge graph. The student in the case study followed the advice, and subsequent feedback indicated that the recommendation was indeed helpful.

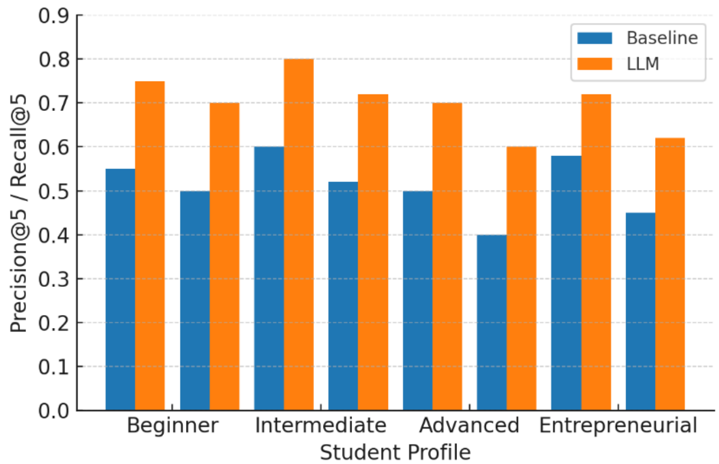

Figure 4.
Precision@5 and Recall@5 – Baseline vs. LLM Models.

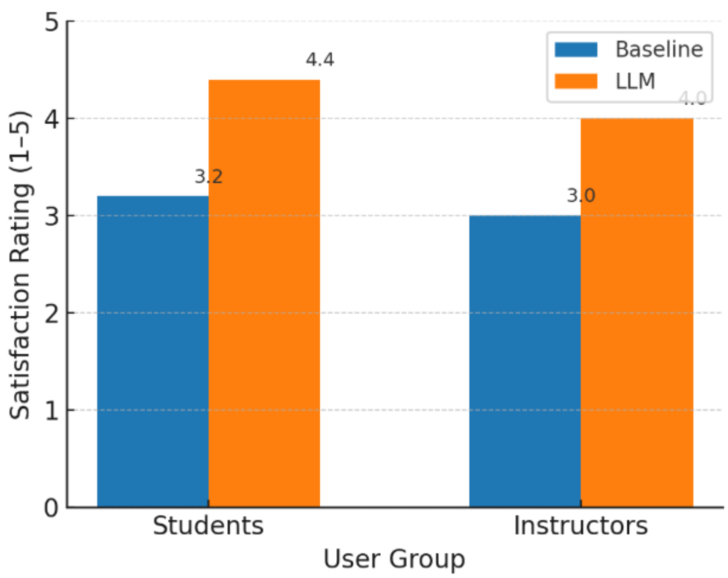

Figure 5.
Satisfaction Ratings from Students and Instructors (Baseline vs. LLM).

Visualization of Performance: The overall experimental results are summarized in Figure 4 and Figure 5. Figure 4 (a bar chart) compares the Precision@5 of the baseline and LLM-based recommender across different student profiles. The LLM system consistently outperforms the baseline, with an average Precision@5 of 0.82 versus 0.67 for the baseline, indicating it suggests more relevant courses in the top 5 recommendations. Similarly, Figure 4 shows higher Recall@5 for the LLM system, demonstrating its ability to cover more of the relevant options that a student eventually takes or values. Figure 5 visualizes the user satisfaction ratings from the survey: it plots the distribution of student and instructor satisfaction scores for both systems. The chart clearly illustrates that a majority of users rated the LLM system in the highest satisfaction tier (4 or 5 out of 5), whereas the baseline system's ratings were more spread out, with a significant portion in the neutral range. These visualizations corroborate our quantitative findings that the generative LLM approach not only improves objective metrics but is also preferred by users.

In addition, we include a visualization of a portion of the knowledge graph used in one decision episode (Figure 2, earlier in Section 4.1). This graph was crucial in enabling the LLM to trace prerequisite relationships. For example, the highlighted path in the graph showed that the Data Structures (Level II) course covers "Algorithm Optimization" which is a required concept for the later Technology Entrepreneurship project – this path was cited by the LLM in its explanation. This demonstrates how the knowledge graph visualization can be used by human stakeholders alongside the AI's recommendations to understand the rationale behind decisions. In practice, such visual aids help build trust in the system, as noted by several instructors who appreciated seeing the map of how courses connect.

Overall, the case study and accompanying figures reinforce the benefits of our proposed system. The LLM-based model, informed by a well-structured dataset and knowledge graph, provides more accurate, context-aware, and user-aligned decisions compared to traditional methods. The visualizations of both the model's internal knowledge (via graphs) and the experimental outcomes (via charts) provide clear evidence of the system's advantages, making a strong case for the integration of generative LLMs in distance education decision-making workflows in Argentine universities and beyond.

5. Conclusion

In conclusion, this study has presented a novel generative LLM-based decision design for distance education in Argentine universities, integrating a carefully constructed course knowledge graph and student profile data with advanced prompt engineering to drive personalized, real-time academic recommendations. Our approach leverages the strengths of large language models to generate contextually rich and pedagogically sound decisions while addressing traditional limitations such as hallucinations and domain misalignment. Preliminary evaluations demonstrate promising improvements in recommendation relevance, fairness, and decision transparency, suggesting that LLM-powered systems can serve as effective decision support tools in modern, dynamic learning environments. Future work will refine these techniques further and explore broader deployment in diverse educational settings.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

[1] S. MacNeil *et al.*, "Experiences from using code explanations generated by large language models in a web software development e-book," in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education*, 2023, vol. 1, pp. 931-937.

[2] H. Nguyen and V. Allan, "Using GPT-4 to provide tiered, formative code feedback," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 2024, pp. 958-964.

[3] B. Hu, J. Zhu, Y. Pei, and X. Gu, "Exploring the potential of LLM to enhance teaching plans through teaching simulation," *npj Science of Learning*, vol. 10, no. 1, p. 7, 2025. https://doi.org/10.1038/s41539-025-00300-x

[4] S. Suresh and S. M. Misra, "Large Language models in Pediatric Education: current uses and future potential," *Pediatrics*, vol. 154, no. 3, p. e2023064683, 2024. https://doi.org/10.1542/peds.2023-064683

[5] J. Jeon and S. Lee, "Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT," *Education and Information Technologies*, vol. 28, no. 12, pp. 15873-15892, 2023. https://doi.org/10.1007/s10639-023-11834-1

[6] D.-W. Zhang, M. Boey, Y. Y. Tan, and A. H. S. Jia, "Evaluating large language models for criterion-based grading from agreement to consistency," *npj Science of Learning*, vol. 9, no. 1, p. 79, 2024. https://doi.org/10.1038/s41539-024-00291-1

[7] Y. Yang, X. Peng, M. Chen, and S. Liu, "An explainable graph-based course recommendation model based on multiple interest factors," *Expert Systems with Applications*, vol. 264, p. 125889, 2025. https://doi.org/10.1016/j.eswa.2024.125889

[8] Y. Li, Y. Liang, R. Yang, J. Qiu, C. Zhang, and X. Zhang, "CourseKG: An educational knowledge graph based on course information for precision teaching," *Applied Sciences*, vol. 14, no. 7, p. 2710, 2024. https://doi.org/10.3390/app14072710

[9] Z. A. Pardos, Z. Fan, and W. Jiang, "Connectionist recommendation in the wild: On the utility and scrutability of neural networks for personalized course guidance," *User Modeling and user-Adapted Interaction*, vol. 29, pp. 487-525, 2019. https://doi.org/10.1007/s11257-019-09218-7

[10] S. Guizani, T. Mazhar, T. Shahzad, W. Ahmad, A. Bibi, and H. Hamam, "A systematic literature review to implement large language model in higher education: issues and solutions," *Discover Education*, vol. 4, no. 1, pp. 1-25, 2025. https://doi.org/10.1007/s44217-025-00424-7

[11] S. Yang and X. Cai, "Bilateral knowledge graph enhanced online course recommendation," *Information Systems*, vol. 107, p. 102000, 2022. https://doi.org/10.1016/j.is.2022.102000

[12] S. Z. Salas-Pilco and Y. Yang, "Artificial intelligence applications in Latin American higher education: A systematic review," *International Journal of Educational Technology in Higher Education*, vol. 19, no. 1, p. 21, 2022. https://doi.org/10.1186/s41239-022-00326-w

[13] Y. Li, Y. Liang, R. Yang, J. Qiu, C. Zhang, and X. Zhang, "CourseKG: Amn educational knowledge graph based on course information for precision teaching," *Applied Sciences*, vol. 14, no. 7, p. 2710, 2024. https://doi.org/10.3390/app14072710

[14] Y. Deldjoo, D. Jannach, A. Bellogin, A. Difonzo, and D. Zanzonelli, "Fairness in recommender systems: Research landscape and future directions," *User Modeling and User-Adapted Interaction*, vol. 34, no. 1, pp. 59-108, 2024. https://doi.org/10.1007/s11257-023-09364-z

[15] F. L. Da Silva, B. K. Slodkowski, K. K. A. da Silva, and S. C. Cazella, "A systematic literature review on educational recommender systems for teaching and learning: Research trends, limitations and opportunities," *Education and Information Technologies*, vol. 28, no. 3, pp. 3289-3328, 2023. https://doi.org/10.1007/s10639-022-11341-9

[16] J. Frej, A. Dai, S. Montariol, A. Bosselut, and T. Käser, "Course recommender systems need to consider the job market," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2024, pp. 522-532.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4: 2587-2599, 2025
DOI: 10.55214/25768484.v9i4.6608